Business Analytics & Data Mining Modeling Using R
Dr. Gaurav Dixit
Department of Management Studies
Indian Institute of Technology, Roorkee

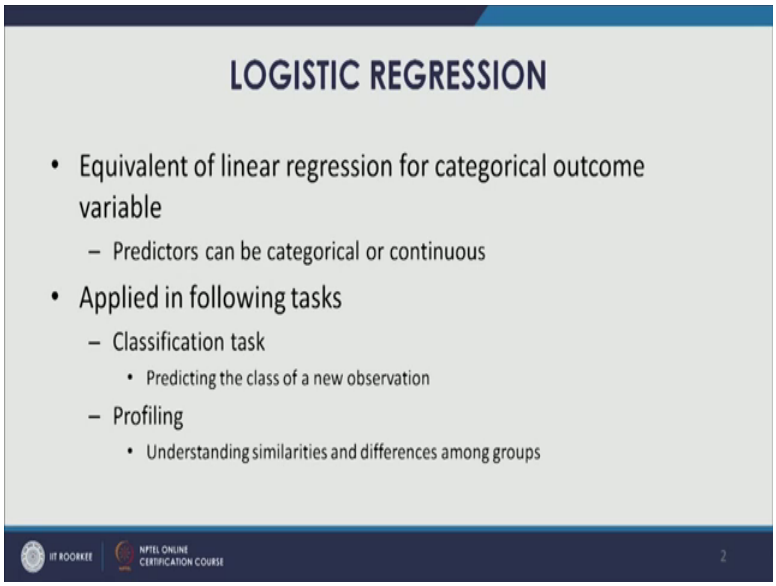
Lecture - 46
Logistic Regression-Part I

Welcome to the course Business Analytics and Data Mining Modeling using R. So, in this particular lecture we will start our discussion on a logistic regression. So, let us start our discussion. So, logistic regression they are equivalent of a linear regression technique that we have gone through in previous lectures, for categorical outcome variable.

So, linear regression that we have discussed before that was mainly for the numeric outcome variable, continuous outcome variable. Now, logistic regression is an is a equivalent is an equivalent technique for categorical outcome variables. So, typically as you would understand that a linear regression is typically used for the prediction tasks and a logistic regression is typically used for the classification task.

However, though the categorical outcome variable we use, the predictors can be categorical or continuous. So, let us understand more about logistic regression.

(Refer Slide Time: 01:28)



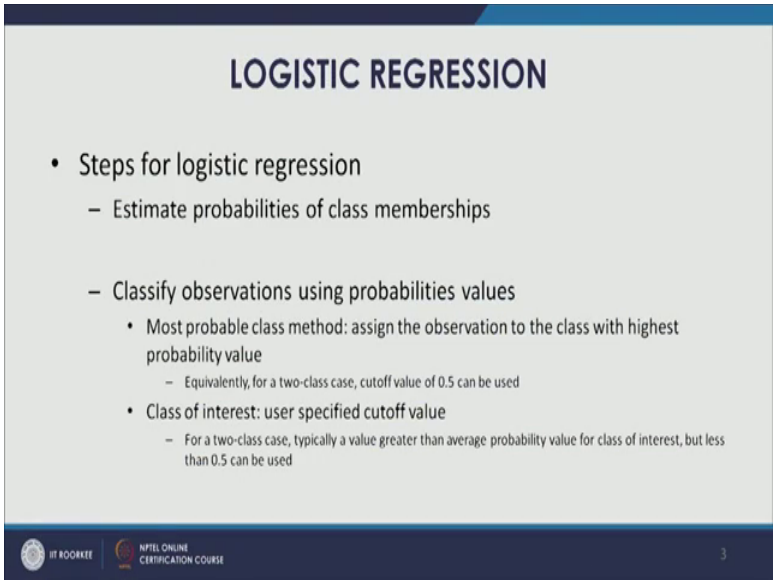
So, typically applied in a following task classification task is one where we predict the class of a new observation. So, that is the you have first category then it this particular

technique could also be used for profiling. For example, understanding similarities and differences among groups right; So, classification task and profiling though the way modeling is done there is not much difference.

So, in terms of modeling these steps for both of both these tasks classification or profiling not much is going to change; however, the ideas the you know objectives are different. So, classifications tasks we would like to predict the class of a new observation in the profiling. We would like to understand similarities and differences among groups.

So, let us move forward. So, what are steps for logistic regression?

(Refer Slide Time: 02:31)



So, first in logistic regression because this is typically for classification tasks we estimate probabilities of class membership. So, if it is a you know m class scenario. So, probably we would like to estimate probabilities of belonging to class one to class two and up to class m. If it is a two class scenario we would like to estimate the probabilities of new observation belonging to a particular observation belonging to class one and class zero right.

So, therefore, with this, so that becomes the first step. So, first we would like to estimate these probabilities, the probabilities of class membership then the second step is about classifying observations using these predicted, using these estimated probabilities value right. So, as we have discussed for other techniques. So, first one is typically the most

probable class method wherein we you know we look at all the estimated probabilities of class membership and we assign the observation to the class with highest probability value.

So, that is the typical scenario where we are looking for you know overall misclassification error, we are looking to minimize overall misclassification error and we do not have any specific class of interest. So, this is the you know typical method that we follow right. This particular most probable class method is equivalent you know for a two class case.

The cutoff value of point five is equivalent to this most probable class method. So, if we have just two classes and we want to apply most probable class method we are just looking to minimize overall misclassification error. In that situation the two class scenario if you know because if the probability of belonging to probability of a record belonging to class one is a you know let us say p.

Then, probability of that particular record belonging to class zero is going to be 1 minus p. So, therefore, you know 0.5 probability value of point half can be used as a as an appropriate cutoff value for a two class scenario and for most problems class method.

Now, as we have talked about when we discussed classification and prediction performance matrix when we discussed different matrix or for classification. We also talked about that we, if you have you know estimated probabilities values like we do in this for this particular technique logistic regression. Then, we can also use a one way axel you know tables to find out the optimized cutoff value for a particular you know for a particular problem for a particular classification problem or task. So, that can also be done right.

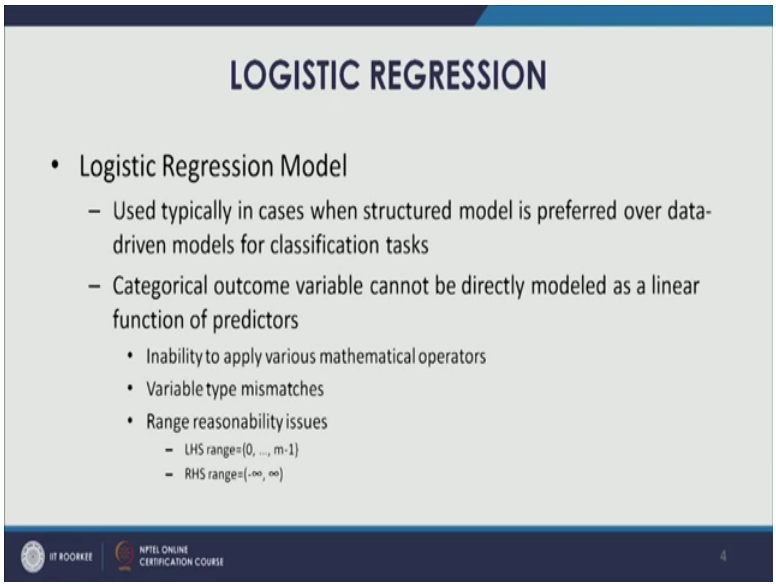
So; however, that might lead to over fitting. So, those things we have already discussed in that particular lecture. Now, the second scenario is class of interest wherein we have the user specified you know we have to use user specified cutoff value. So, in the class of interest we might have one particular class which is which is a low probability event.

So, therefore, we would like to identify you know more members of that particular class and. So, we have talked about different scenarios using cost based method, misclassification cost and other things in previous lecture already. Right now, for class of

interest, if we are dealing with a two class case. So, typically a value that is a greater than you know average probability value for class of interest, but less than 0.5 probably that can be used. Still just a kind of rule of thumb based on the, you know experience in modeling.

So; however, depending on the depending on the probabilities value and depending on the you know we have that cost matrix we can find out a find out a good enough cutoff value or optimized cutoff value in class of interest scenario. So, these are two typical step; first we would like to estimate probabilities and then these probabilities are used to classify observations. Now, let us move forward.

(Refer Slide Time: 07:02)



So, now, I will discuss more about the logistic regression model as such. So, typically this is this particular model. So, as we talked about this is for classification task, so this particular model is typically used in cases where a structure model is preferred over data driven models right. So, we have a discussed a few data driven model right like new base and K N N right.

So, we have discussed a few data driven models. However, when we required a structural model when we have some assumed functional form between the outcome variable. And you know you know and the set of predictors or we understand that there is some sort of relationship between the predictors and outcome variable. And if we are looking to structure that we are looking to model that then probably logistic regression is an is a is

an equivalent technique for the same you know a structural technique for classification task just like linear regression is for prediction tasks.

So, as we understand that categorical outcome variable if we try to model it as a linear function of predictors. So, that is not possible because of various reasons. So, as mentioned in the second point here. Categorical outcome variable cannot be directly modeled as a linear function of predictors right because on when you have a Y a categorical predictor so you are going to have you know a few values that are nothing, but the, but the different categories. Let us say if there are m classes so they are going to be if we express that in numerical code format then it is going to be 0 or 1 or 2 or 3 up to m minus 1.

So, those could be the classes. So, now, if you try to you know model that as if a linear function of predictors where predators could be continuous and categorical. So, that range would be minus infinite to an infinite.

So, that there are going to be certain problems. So, which are expressed some of them, some of these issues are expressed in this particular slide as well. Inability to apply various mathematical operators one so because your, categorical variable could be nominal variable. So, more often not the categorical outcome variable that we might use it could be nominal variable.

So, therefore, we have already discussed in one of the lecture the kind of you know operators that can be used. So, equal to n is not equal to. So, these are the you know two operations that can be performed. And therefore, the you know that linear modeling cannot be applied.

Now, variable type mismatch that is also obvious that, in the outcome variable we have just that is categorical and in the you know set of predictors where we might have continuous and categorical variables there. So, therefore, interpretation and modeling and everything is mismatch there.

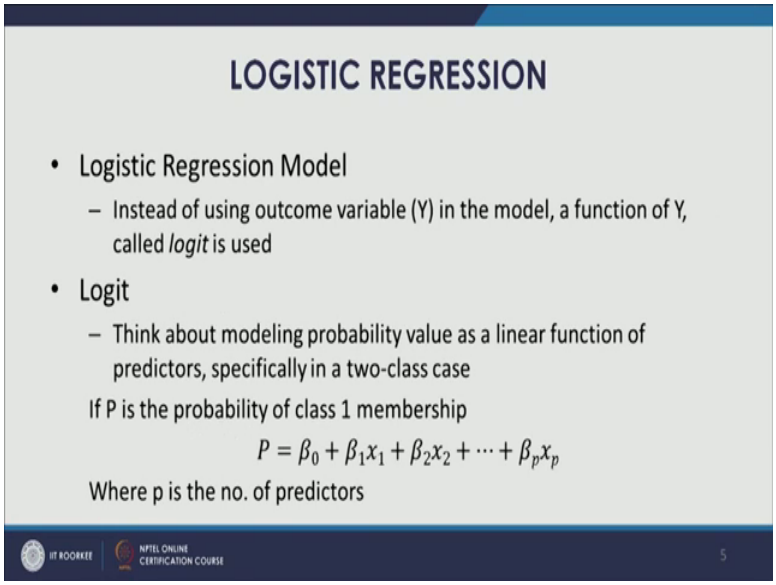
Now, there another thing to understand is range reasonability issues. For example, categorical outcome variable as I said that the values could be one of you know if we are expressing the categorical variable numeric code format. Where, 0 is representing you know one of the class 0 then 1, class 1 and similarly class m is being represented by m

minus 1. So, in that fashion you will have one of these values from 0 to m minus 1 in LHS side that is outcome variable side.

And then RHS side you would have the predictors and therefore, some of those predictors could be numeric variables as well. So, therefore, the range could be minus and finite to infinite.

So, therefore, that range reasonability issues would also occur. So, lots of things, very you know few that is just equal to n is not equal to operators can be applied. Variable type mismatch is there reason range reasonable reasonability issues are there. So, because of this categorical variable cannot be directly modeled as a linear function of predictors. So, let us move forward.

(Refer Slide Time: 11:20)



So, how do we, so what we do in logistic regression model then? So, instead of using outcome variable Y which is typically, which is actually the categorical variable as well; So, instead of using this categorical outcome variable Y in the model if function of Y called logit is used.

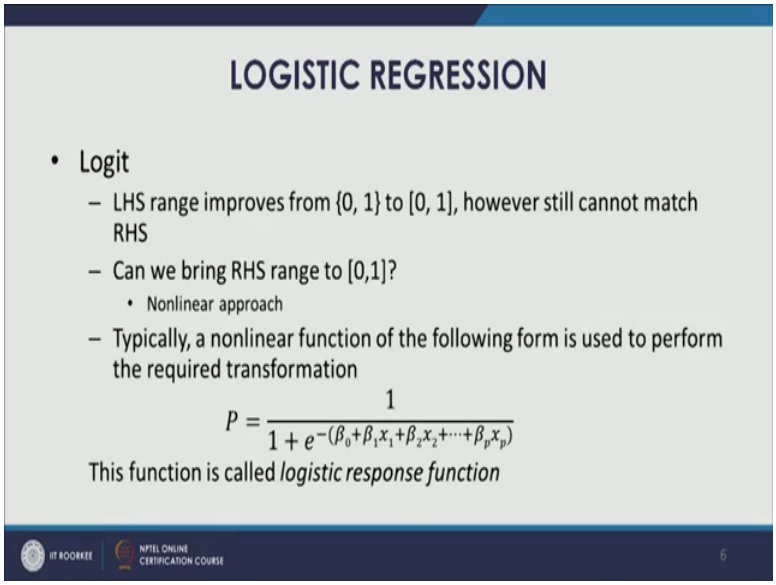
So, let us understand what this function is about. So, before we reach to our standard logistic regression model and discuss logit and other related concepts. Let us think about a modeling probability value as a linear function of predictors specifically in a two class scenario right. So, if we have just two class as I mentioned in the in the in this slide as

well we are in a two class scenario. So, we just need to you know estimate one probability values that is probability of belonging to class 1.

The other would be 1 minus of this particular probability value. So, therefore, the probability value P can be expressed in this form as a linear function or predictors beta 0 plus beta 1 x 1 plus beta 2 x 2 plus beta p x p, if there are p small p that is p number of predictors.

So, what is the problem with this particular formulation? So, let us discuss this. So, if we have a two class scenario and we express the probability value of class one membership as a set linear function operators then, what are going to be problem with this? So, let us look at the LHS range now.

(Refer Slide Time: 12:56)



At LHS range will improve from a set of two values that is a 0 and 1 to you know this range from 0 to 1 right. So, now because this is LHS is now probability. So, therefore, the probability value can range from 0 to 1. So, the LHS range improves from just two distinct values to is small you know range that is between 0 and 1. However, it still cannot match RHS range where is minus and finite to infinite.

So, what we can do? Can we bring RHS range? That is right side range to 0, 1 level. So, it is possible with some transformation. So, this is typically non linear approach we have to take. Some non linear transformation we would have to perform in a manner that the

expression that we have on the right hand side that you know linear that linear function of predictors that we have on the right side as expressed here; In a form where it starts taking in a non linear form where, it start taking value values from 0 to 1.

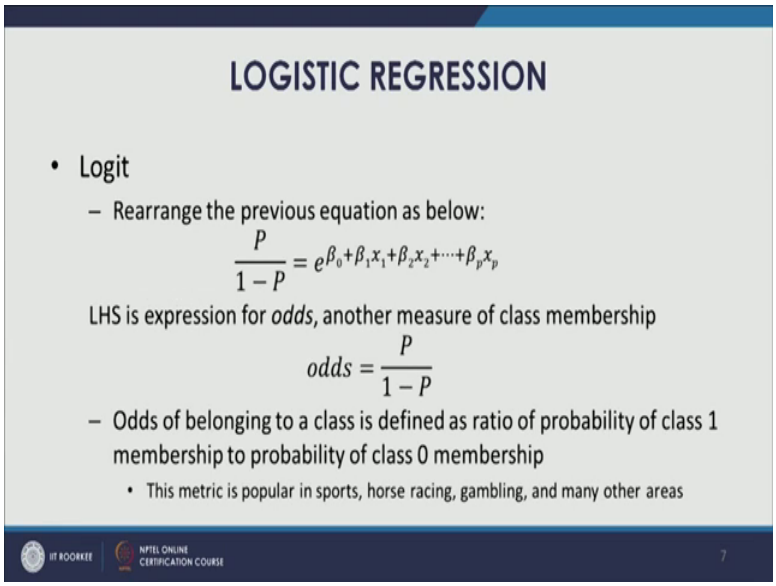
So, typically as you can see in the slide a non linear function of this form which is called logistic response function can actually be used to perform this transformation. So, we can express probability value as this probability value P it can be expressed as 1 divided by 1 plus exponential of minus of beta 0 plus beta 1 x 1 plus beta 2 x 2 plus up to beta p x p, that is a linear function of our predictor. So, you can see from here that the RHS side, the earlier RHS side that we had now it is you know it is within this exponential you know this is power to the exponential negative power to the exponential.

And now irrespective of the values, that is being taken by with this linear form of predictors the value of right hand side would be would be would be in that 0, 1 range. And now you would see that left hand side and right hand side both are in the same range 0 to 1.

Right; So, left hand side is probability values probability value in two class case or in a for that matter any class case this is going to be 0 and 1, between 0 and 1 and right hand side is also now between 0, 1. As you can see when denominator it you know in this exponential function it can it can range from 0 to infinite. So, therefore, when it becomes when it is approached 0; So, the value is going to be 1; when it approaches infinite then the value is going to be close to 0. So, therefore, this right hand expression is going to be in that range 0 to 1.

Now, once these two ranges once range this ability is you know now we can move forward.

(Refer Slide Time: 16:02)



So, the same, the previous equation that we just saw can be rearranged in this form. So, now, instead of you know having that form we can use this particular rearrangement wherein we have on LHS side P divided by 1 minus P and on the RHS side we have the exponential of beta 0 plus beta 1 x 1 plus beta 2 x 2 up to beta p x p. So, in this particular formulation you would see that now the range here has changed, but it is now it been you know 0 to infinite for both the side LHS and RHS as well.

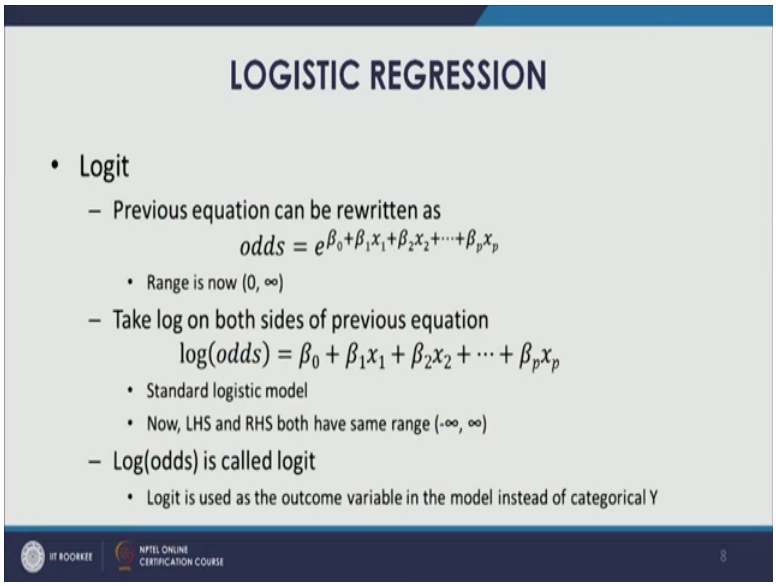
Now, if we look at this particular equation in the RHS side, now this is more of a proportional form. So, if we use this particular model you know the interpolation would be in percentage terms right. However, we can do a certain more transformation to bring it to typical standardized form of equivalent form of a linear regression; So, that being the objective.

So, we would like to reach to the equivalent you know standard formulation that we use in linear regression for the logistic regression as well. So, now, LHS the expression that we see P divided by 1 minus P that is actually nothing, but the definition of odds; So, odds is another measure of class membership where, it can be defined as expressed as odds equal to P divided by 1 minus P. And it can be defined as odds are belonging to a class is defined as ratio of probability of class 1 membership to probability of class 0 membership right.

So, it is a ratio of you know whether there is going to be success or failure. So, this particular now we can use instead of using this expression P divided by 1 minus P we can replace it with this odds which is also another measure of class membership.

So, this particular metric odds is popular in sports, horse racing, gambling and many other areas. So, now, you can also understand that range for odds metric is from 0 to infinite. So, the same point is here. So, in previous equation now it can be rewritten as odds equal to exponential e to the power beta 0 plus beta 1 x 1 plus beta 2 x 2 up to beta p x p. So, a range is now 0 to infinity as you can see both LHS and RHS. We have already matched these ranges and that same thing continuous.

(Refer Slide Time: 18:46)



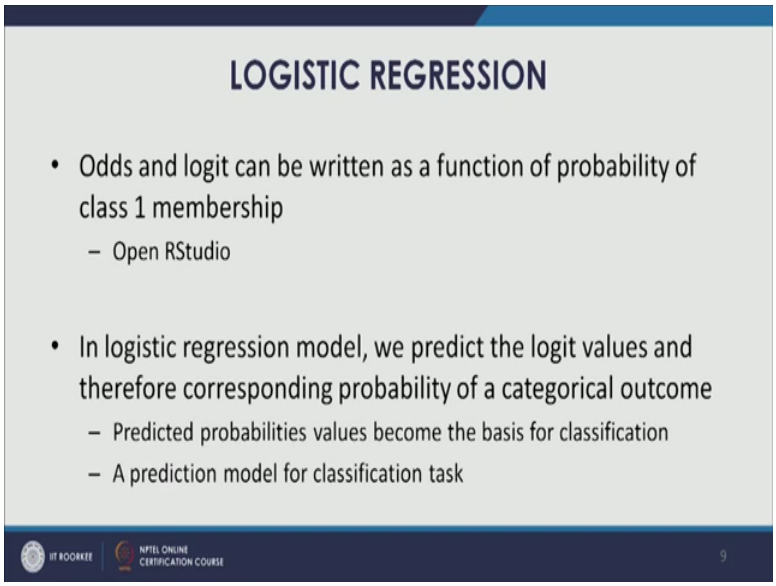
Now, if we take a log on both sides of this particular equation then, what we get is log of odds and on the right side we get a linear function of predictors that we wanted to achieve. That is beta 0 plus beta 1 x 1 plus beta 2 x 2 up to beta p x p. Now once we take log now let us look at the you know arrange reasonability on first look at let us look at the LHS side log odds. So, as we know that log of a log of a value you know that log based function they will take range from minus infinite to infinite. And RHS is now in the linear function format. So, it is it will also take the same range minus infinite to infinite.

So, range is matching and this particular formulation that we now see log of odds being expressed as a function of these predictors beta 0 linear function of these predictors; that

is beta 0 plus beta 1 x 1. So, this is the standard logistic model. So, this is a standard formulation of a logistic model.

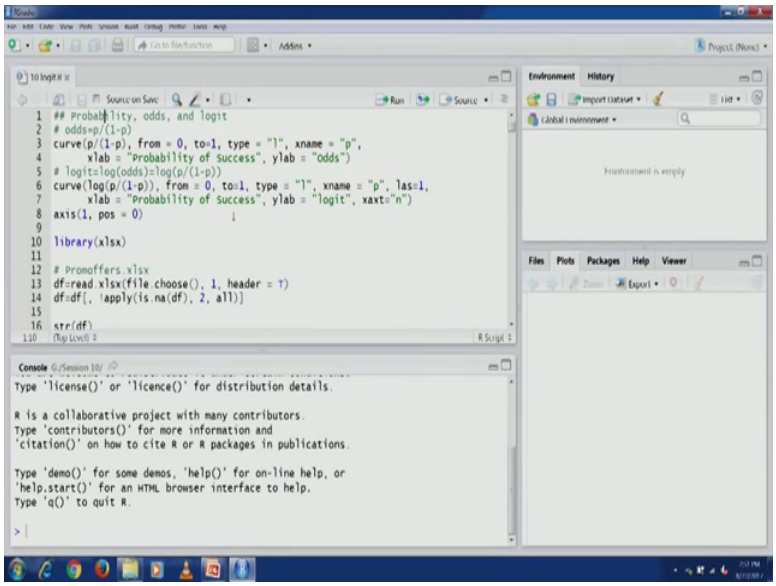
So, this is the formulation that we typically used in a linear regression modeling. Now this particular term on the this particular term in the left hand side log odds is called logit. So, as we talk about as we talked about that instead of using the categorical outcome variable, we would be using logit. So, this is how this is the formulation this is expression this is the expression for logit. So, logit is nothing, but log odds which is log P divided by 1 minus P. So, this particular logit is going to be used as the outcome variable in the model instead of the categorical Y.

(Refer Slide Time: 20:43)



Now, odds and logit as we have understood and now both these matrix can actually be written as a function of probability of class one membership right. To understand more about the relationship between these two these three matrix let us see few plots in R studio. Let us look an R studio.

(Refer Slide Time: 21:12)



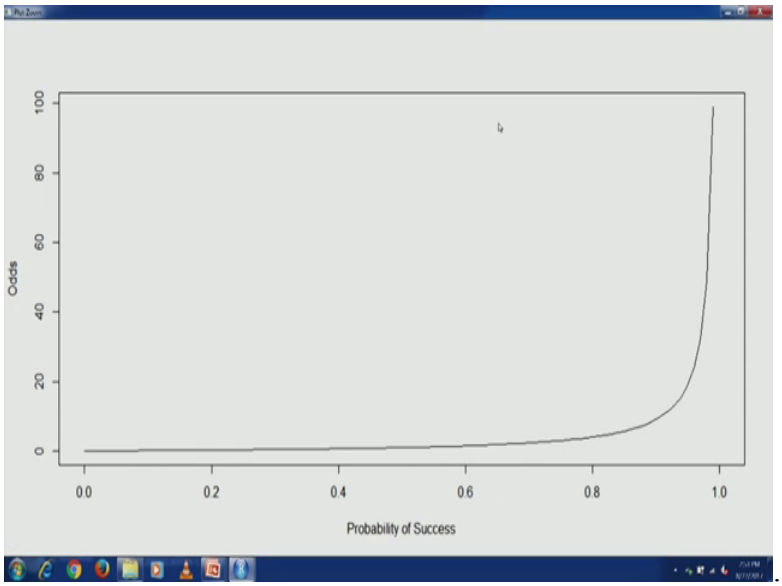
So, what we will do we will look at we will create some you know some plot for this. So, odds versus probability value of class 1 membership and also logit versus probability of class 1 membership.

So, we will try to understand the range reasonability as well as a will understand how the values are been taken. So, odds as we talked about, odds is can be expressed as P divided by 1 minus P. So, this is the particular function that can be used to plot this particular mathematical relationship where, curve is the function. So, first argument is the expression. So, in this case probability value; So, P divided by 1 minus p.

So, this is the expression that we want to plot right. So, P divided by 1 minus P, the values can range from 0 to 1 right because the probabilities value. So, that is the range for a probability value. So, the same has been specified in argument second and third. So, third is from 0 to second argument is from 0 and third argument is to 1. Now, type of the plot is going to be l, that is linear you know linear plotting. And then this is expression that x name this particular argument is p because it is the value of p that we are changing from 0 to 1.

So, a label for x axis is going to be probability of success because this particular these particular values are probability of class 1 membership and the label for y axis is odds. So, let us execute this particular code and we would see a plot here, let us zoom in.

(Refer Slide Time: 22:57)



Now, you would see that on x axis we have the values ranging from 0 to 1 and on y axis we have the values ranging from 0 0 to infinite. However, we are just showing up to 0 to 100. So, on y axis we have odd. So, we can see as the values increase from you know left to right on x axis from 0 to 1 the value of odds keeps you know keeps on increasing all right.

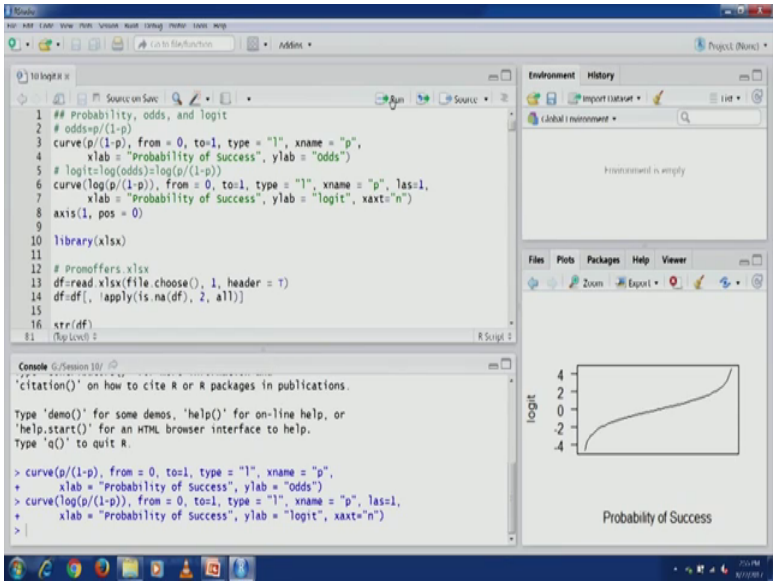
So, this particular odds value will keep on increasing. So, as we approach you know as you can see 0.8. So, this up to this 0.8, there is sharp increase in the odds values and you know when we reach on x axis when, we reach to 1 the odds value on a y axis it reaches to infinite. So, this is the plot for you know odds.

So, if we are talking about you know typical default cutoff value that we use is 0.5 then the corresponding odds value is going to be you know somewhere around this particular you know 0.5 5value.

So, the corresponding odds value is going to be 1 right. So, let us move forward. So, similarly we can express we can generate a plot for logit versus p that is probability of class 1. So, logit can be expressed as log odds and which is nothing, but log p divided by one minus p. So, same thing we can express in the this particular function curve where, first argument is going to be log p divided by 1 minus p. Second and third arguments are going to be remain same from 0 to 1 because probability values will range from 0 to 1. Type of plot is again l name of this name of x is p. This p is the that that variable that we

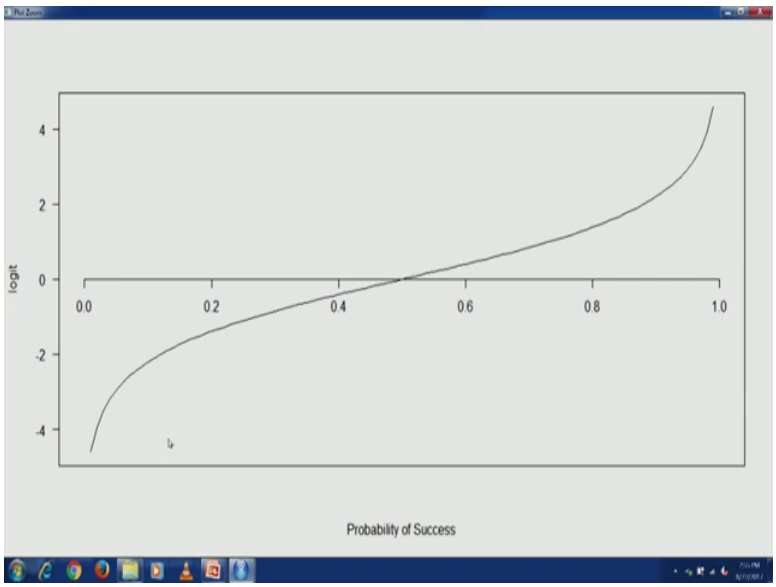
have and this is styling detail then, we have appropriately we have specified the labels for x axis and y axis.

(Refer Slide Time: 25:02)



Let us create this plot as well. So let us also create x axis, you can see in the curve function we have last argument is x a x t as n; that means, x axis was not plotted. Now we will create x axis separately. Let us do this, now let us zoom into the plot.

(Refer Slide Time: 25:23)



Now, from here you can see the values along x axis range from 0 to 1 for this particular plot. And you would see that on y axis we have logit and the values range from minus

infinite to infinite. However, for this for this particular plot we see typically the values from minus 4 to 4. Now as we you know move from a left to right along x axis. So, as the probability values increase from 0 towards 1, you would see up to you know 0 to 0.5. So, 0.5 this particular you know logit value becomes you know 0 that is because this particular expression is for log p divided by 1 minus p. So, p divided by 1 minus p, will be oh 1 for a value of p, even value of p is 0.5.

So, therefore, log of 1 is going to be 0. So, therefore, at this probability value p equal to 0.5 this particular logit function is 0 right. And you can see from when the probability values are near to close to 0, then the logit values are closer to minus infinite. They are high values you know on the in the in the negative sign negative side and as we you know a increase values from 0 to 1 from 0.5 to 1 the logit values keep on keeps on increasing and as we approach 1 we approach, we consider values close to 1, then the logit values they approach infinite.

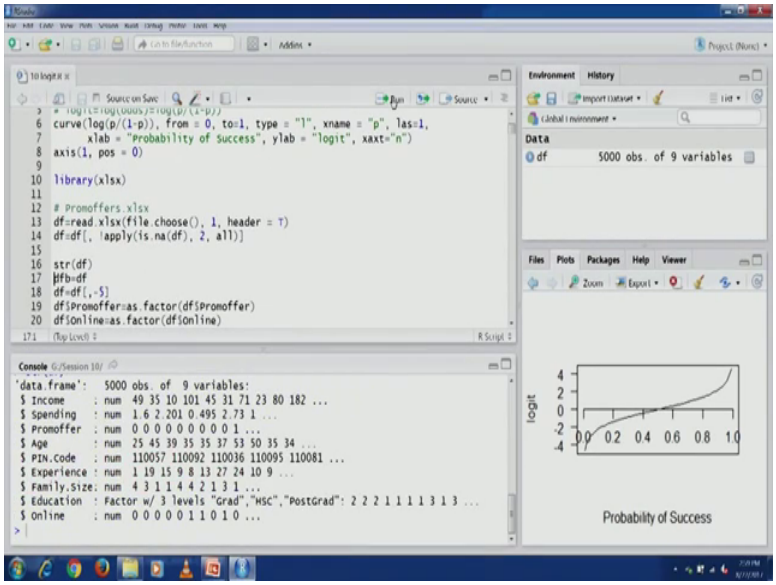
So, the range for logit is minus infinite to infinite and you can see the probabilities range values then 0 to 1. From this so this is the relationship between odds and odds logit and probability value. So, from this we can understand that the standard formulation for logistic model that we just saw once we estimate those beta values, it is always possible for us to compute the you know corresponding probability value from a logit value.

So, with this we will do one exercise in also the data set for a logit model, that we are going to develop is a this particular data set we are going to use promotional offers. So, this particular data set we have used in previous technique as well classification and regression trees. So, what we are going to do is we are already familiar with the data set and variables what we will do? We will first start with the building a simple a logistic model just like simple linear regression where we just regress outcome variable with one particular predictor.

So, in this perfect case, the promotional offer will build an array between promotional offer and income variable. So, let us first load this package x plus x let us import the data set. So, promotional offers data set that we have already seen in previous lecture, previous technique and that lectures as well. So, we have 5000 observations and this is about we are we can build a classification model of where in a particular customer or (Refer Time: 28:46) whether they are going to accept or reject a promotional offer. So,

now, using logistic radiation we will try and build a classification model. So, you can see that data set is now imported in our environment.

(Refer Slide Time: 29:02)



So, let us look at the apply structure function. So, we can see these variables income is spending promotional offer. So, first we would like to start building you know simple logistic model logistic regression model between promotional offers and income; Income being the predictor and promotional offer being the outcome variable.

Other variables also we are familiar with age, pin code, experience, family size, education, and online. So, first let us take a backup of this particular data frame. Now, in this particular logistic regression excide you would not like to include pin code because as we have understood too many categories we will have to deal with this. So, as of now will not like to consider for logistic regression this particular variable.

So, let us create make that change in the data frame. Now two variables promotion offer that is our outcome variable, let us make it factor variable and also the online that is the whether the particular customer is online active or not. So, let us change these two variables. Now let us look at this structure, now all the variables are in the appropriate format variable type income spending numeric and then promotional offer is factor, then age and experience and family size.

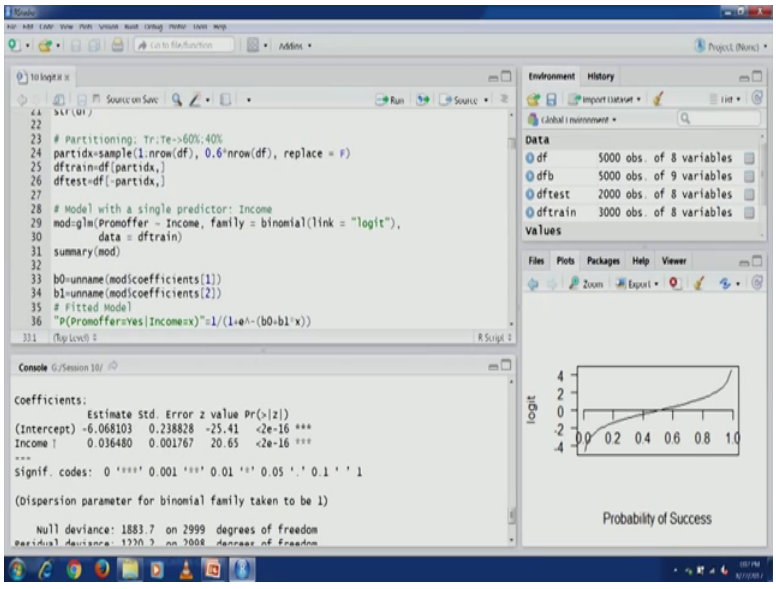
They are numeric, then education and online. So, they are factor. So, as we have been doing in other techniques as well first let us start with partitioning. So, we will go for 60 percent observation in training partition and 40 percent observation in test partition. So, 60, 40 that is the proportion will go for. So, let us create the partitioning.

now to build the model. So, as we said we will just build a model between these two variables promotional, offer and income just to one just a single predictor. So, GLM is the function that is used to build logistic regression model. So, GLM first in GLM function first we will have to pass the formula for our model, that is promotional offer till day income and then second argument is about the family that we have to pick.

Families actually you know about telling this particular function that we would like to build logistic regression model. So, you can see family binomial. So, logistic is part of the binomial family. So, link you can see logit. So, we would like to build logit model and the data is that training partition. So, let us run this code in the logistic model.

Let us look at the summary. So, in this summary you would see that we have an intercept term, we have income term estimate for these two.

(Refer Slide Time: 31:34)



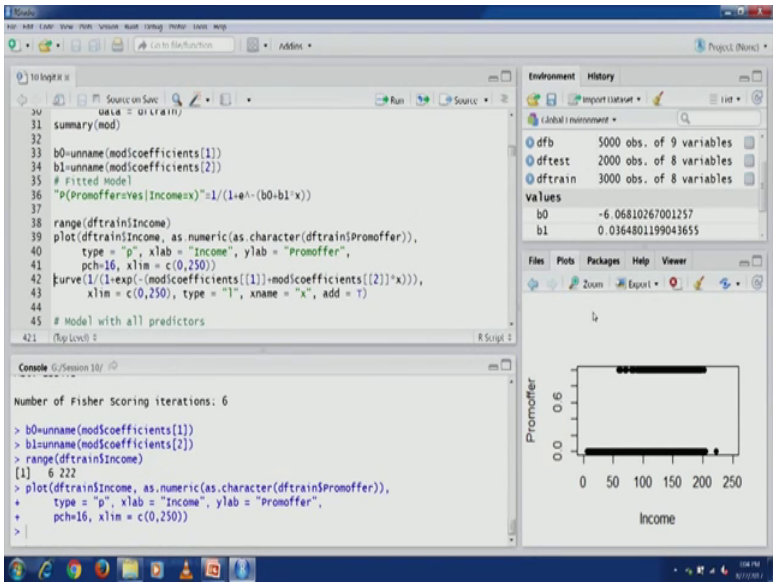
So, beta 0 and beta 1, beta 1 for income that we can see and we can see both of these you know estimates are significant right. So, beta one for income that is also significant here as you can see; Now, if we want to express this in terms of for final model then we

would have to extract these beta 0 and beta 1 value. So, this is how we can extract. So, model object that we have just created, it will have this attribute coefficient and within the coefficient we will have just two values. We can extract using in this fashion.

Unnamed function would just remove the names for this particular attribute, this particular coefficient attribute. So, we are not interested in the name. So, we just are not interested in extracting the values; so, that we can write our formulation of the final model of logistic regression. So, let us introduce these two commands. So, the fitted model, we can write in this fashion. Probability of a particular observation you know accepting a particular individual, accepting particular customer, accepting promotional offer. Given their income label as x, can be expressed in this form; e 1 divided by 1 plus e to the power minus b 0 plus b 1 into x.

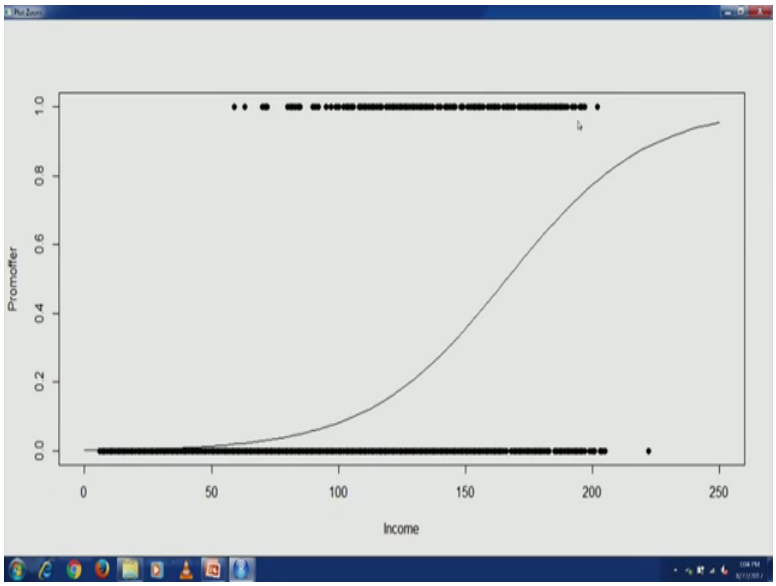
So, this particular expression one single predictor model fitted model can be expressed in this form. If you want we can plot this. So, let us look at the range of income variable and create a plot. So, the plot is between income variable that is the predictor in this case and the proportional offer and the promotional offer that is the outcome variable. So, let us create this plot.

(Refer Slide Time: 33:43)



So, we would see that now let us add the this curve as well this curve that the model that we have just fitted. So, you can see using the curve function we can express the same model here and then we can add it to the plot.

(Refer Slide Time: 34:07)



So, this is the plot for our logistic response model and all the you know dots that you see these are our observations right and. So, this is our fitted model here. So, as you can see income values they range from 0 to 50 for you know that this is the scale and promotional offer that is nothing, but the probabilities value 0 to 1.

So, at this point, we will stop here and more on logistic regression and this particular plot as well we will discuss in the next lecture.

Thank you.